B7

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/663,812	09/15/2000	Julian C. Chen	YOR9-2000-0144US1 (8728-	3772

7590 12/27/2004

Frank Chau Esq
F Chau & Associates LLP
Suite 501
1900 Hempstead Turnpike
East Meadow, NY 11554

EXAMINER	
ALBERTALLI, BRIAN LOUIS	

ART UNIT	PAPER NUMBER
2655	

DATE MAILED: 12/27/2004

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	09/663,812	CHEN ET AL.	
	Examiner	Art Unit	
	Brian L Albertalli	2655	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☐ Responsive to communication(s) filed on _____.

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) 1-21 is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) 1-21 is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☐ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: _____.

DETAILED ACTION

Response to Arguments

1. Applicant's arguments filed September 9, 2004 have been fully considered but they are not persuasive.

2. In the previous Office Action, the term "semantic units" was interpreted as being a letter, phoneme, syllable, morpheme, word, or phrase. The term "semantic units of words", as presented in the Applicants amended claims, could indeed narrow the definition of "semantic units" relied upon in the previous rejection. However, lacking any explicit definition thereof in the specification, the term "semantic units of words" still encompasses *any unit of semantics of a word*. The specification simply states that syllables and morphemes are *examples* of "semantic units" without giving an explicit definition. Similarly, in the arguments presented, "semantic units of words, *such as* syllables or morphemes" (emphasis added) does not clearly limit the term "semantic units of words" to only syllables and morphemes. Therefore, the term "semantic units of words", as interpreted herein, legitimately includes letters, phonemes, syllables, morphemes, or the complete word.

3. In response to applicant's argument that the references fail to show certain features of applicant's invention, it is noted that the features upon which applicant relies (i.e., transcribing textual data into semantic units of words implies understanding of the language, page 7, line 22 to page 8, line 6) are not recited in the rejected claim(s).

Although the claims are interpreted in light of the specification, limitations from the specification are not read into the claims. See *In re Van Geuns*, 988 F.2d 1181, 26 USPQ2d 1057 (Fed. Cir. 1993). Furthermore, Nanjo et al. states that the method requires "no *special* understanding of the language being indexed or searched" (column 3, lines 12-15). To state, then, that Nanjo et al. "does not even require *an* understanding of the language being indexed" (arguments, page 8, lines 3-4) implies that Nanjo et al. does not disclose any understanding of the language being indexed whatsoever. However, since Nanjo et al. discloses the method requires no *special* understanding of the language being indexed, it could be reasonably argued that Nanjo et al. discloses at least a rudimentary understanding of the language being indexed.

Nanjo et al. discloses the textual data (Fig. 3, document 321 and document 322) are text files (*.txt). A text file must have been transcribed, either through manual typing, speech recognition, character recognition, or some other transcription method, from textual data in order to be stored as a text file. The textual string includes a Kanji string A (KKKKK, 331), each individual Kanji character having been stored in a separate position in the text file (see 350, positions 0-4). A single Kanji character is a semantic unit of a word, which can represent either a syllable or a morpheme or the entire word. Nanjo et al., therefore, does disclose transcribing data into semantic units of words.

4. Furthermore, the argument (page 8, lines 7-13) that the indexing involves *breaking a string into preliminary index terms, which are each a longest substring that contains only word characters*, is irrelevant because it only applies to the creation of

preliminary index terms, as illustrated in steps 515 and 521 in Fig. 5. The final index is not completed until after the completion of step 600 in Fig. 5. In Fig. 6, the creation of the final index comprises a step 700 to create a step index of the Kanji characters. In Fig. 7, the string of Kanji characters is broken into smaller tokens, with *each token being saved as an index term* (column 14, lines 40-41). Each token of Kanji characters can represent semantic unit such as a word, or morpheme, or in the case of the single character "g" a syllable.

Nanjo et al., therefore, does disclose transcribing/indexing based on semantic units of words, even though without a special understanding of the language being indexed.

5. Therefore, the claim rejections made in the previous Office Action stand.

Claim Rejections - 35 USC § 102

6. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

7. Claims 1-3, 8, 11-12, and 15 are rejected under 35 U.S.C. 102(b) as being anticipated by Nanjo et al. (U.S. Patent 5,778,361).

In regard to claim 1 and 15, Nanjo et al. discloses a method and a program storage device (404) for performing the method steps that is used for managing (indexing) a textual database.

The textual data (Fig. 3, document 321 and document 322) are disclosed as text files (*.txt) that inherently must have been transcribed into semantic units before indexing (column 9, lines 7-11).

The textual data is stored in a textual database (computer based collection of documents) (column 3, lines 7-10).

An index of the textual database is created based on semantic units. As broadly recited in the claim, the term "semantic units" has been interpreted as being any unit of semantics, i.e. a letter, phoneme, syllable, morpheme, word, or phrase. The textual data is indexed by an inverted list (302) that stores the semantic units (indexing terms) and references the documents containing each term (column 9, lines 11-15).

In regard to claim 2 and 3, Nanjo et al. discloses that the semantic units used for indexing are both syllables and morphemes. In Fig. 7, a flowchart (700) is given that shows the steps for indexing input characters. The code in Fig. 7 breaks a string of Japanese kanji characters into substrings until the final semantic unit (index term) is the final kanji character of the preliminary index term (column 14, lines 26-41 and column 15, lines 13-17). A single kanji character can be both a morpheme or a single syllable or both.

In regard to claim 8, Nanjo et al. discloses the index is a hierarchical index in which a semantic unit (indexing term) points to another mode of data. The index is an inverted list (302) with nodes (307 and 308) that point to a leaf structure (309-312). The leaf structures point to a centralized list of documents (column 9, lines 32-43).

In regard to claim 11, Nanjo et al. discloses searching the textual database for target textual data (query entered into search string box Fig. 2, 203) using the semantic unit index (content index) (column 15, lines 31-38 and column 16, lines 41-43).

In regard to claim 12, Nanjo et al. discloses converting a target word into a string of semantic units to perform the searching step. In Fig. 8, a flow diagram (800) is given in which a target word (string of characters to be searched) is entered (803) (column 15, lines 44-45). The target word (string of characters) is converted into semantic units (search terms) by checking whether the current character is a separator (811), or a character type transition (819). If either of these conditions is met, a key offset (KO) and key limiter (KL) type are entered into a key buffer as a delimiting semantic unit (search term) (column 16, lines 1-19 and column 12, lines 43-48). Once the target word (string of characters to be searched) has been converted to semantic units (search terms), a search is conducted on those semantic units (900).

Claim Rejections - 35 USC § 103

8. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

9. Claims 7, 10, and 14 are rejected under 35 U.S.C. 103(a) as being unpatentable over Nanjo et al.

In regard to claim 7, Nanjo et al. discloses all features of the instant claimed invention except transcribing using semantic-unit based stenography. Semantic-unit based stenography, as broadly recited in the claim, has been interpreted as an input device based on semantic units (e.g. letters). A keyboard is an input device based on semantic units and, as is well known in the art, can be used to transcribe textual data. Using a keyboard to transcribe data would provide an additional transcribing method that could be used if other transcribing methods, such as speech recognition or character recognition, were not accurate enough. It would have been obvious to one of ordinary skill in the art at the time of invention to modify Nanjo et al. so that semantic-unit based stenography was used for transcription in order to provide an additional, more accurate transcribing method.

In regard to claim 10, Nanjo et al. discloses that language symbols can be represented by series of bits. Nanjo et al. further discloses that given a particular

coding scheme, a table can be constructed that translates a given code into an appropriate character of the language (column 2, lines 12-15).

Nanjo et al. does not specifically disclose using the table to convert the index of textual data into a universal index. Converting the index of the textual data into a universal index would allow the textual data to be searched in any language.

It would have been obvious to one of ordinary skill in the art at the time of invention to convert the index of the textual data into a universal index so that the textual data could be searched in any language.

In regard to claim 14, Nanjo et al. discloses that the target textual data (the list of objects that satisfy the search criteria) is displayed (Fig. 8, 837) (column 16, lines 32-35).

Nanjo et al. does not disclose displaying one semantic unit forward and backward for a given length based on a user request.

It would have been an obvious matter of design choice to modify Nanjo et al. so that in addition to the target textual data being displayed, one semantic unit forward and backward was also displayed, since the applicant has not disclosed that displaying one semantic unit backward and forward solves any stated problem and it appears the display function would work equally well with any number of semantic units displayed on either side of the target textual data to provide needed context.

10. Claims 4-6, 16, and 18-21 rejected under 35 U.S.C. 103(a) as being unpatentable over Nanjo et al. in view of Holt et al. (U.S. Patent 5,960,447).

In regard to claims 4-6, Nanjo et al. discloses all the features of the instant claimed invention except associating the textual data with audio data wherein the step of indexing further comprises indexing the audio data with semantic units, time stamping the semantic units, and decoding the textual data with a recognition system utilizing a language model based on semantic units.

Holt et al. discloses a tagging and editing system that links textual data (word processor file Fig. 2, 60) to an audio file (53). Each semantic unit (word) in the textual data (word processor file 60) is indexed in the audio file (column 4, lines 1-18). The semantic units (words) are time-stamped (a time code pointing to a particular starting point in the audio file) (column 4, lines 5-7). A recognition system (52) receives speech as an input from the microphone (50) and transcribes the speech to textual data (text words) (column 3, lines 16-20). A speech recognition system typically utilizes a language model based on semantic units (e.g. phonemes in a HMM word model).

Adding indexes to textual data transcribed with a recognition system corresponding audio to data that is time-stamped, as taught by Holt et al., to a system of managing a textual database, as taught by Nanjo et al., would allow the playback of associated audio for each recognized semantic unit, thereby helping in correction and proof reading of a textual database, as taught by Holt et al. (column 4, lines 29-31).

It would have been obvious to one of ordinary skill in the art at the time of invention to add time-stamped indexes to audio data corresponding to the textual data in order to help in the correction and proofreading of a textual database.

In regard to claim 16, Nanjo et al. discloses a system for managing (indexing) a textual database (400) that includes a textual database (objects 406, 407, 408, and 409) and an index generator (index program 415) that generates an index based on semantic units, which indexes the textual database with the corresponding semantic units (content index 410) (column 11, lines 38-66 and column 12, lines 1-6).

Nanjo et al. does not disclose a recognition system for transcribing textual data into corresponding semantic units.

Holt et al. discloses a recognition system (52) for transcribing textual data into corresponding semantic units (words). Combining the system for managing a textual database as taught by Nanjo et al. with a recognition system as taught by Holt et al. would allow a user to transcribe textual data without having to use a keyboard.

It would have been obvious to one of ordinary skill in the art at the time of invention to modify Nanjo et al. to include a recognition device in order to transcribe textual data without having to use a keyboard.

In regard to claim 18, Nanjo et al. does not disclose a language model is based on semantic units.

Holt et al. discloses a speech recognition system (52) to transcribe textual data. It is well known in the art that a speech recognition system typically utilizes a language model based on semantic units (e.g. phonemes in a HMM word model), so using a language model based on semantic units to transcribe textual data would be obvious to one of ordinary skill in the art at the time of invention in order to increase the amount of correctly recognized data.

In regard to claim 19, Nanjo et al. discloses that language symbols can be represented by series of bits. Nanjo et al. further discloses that given a particular coding scheme, a table can be constructed that translates a given code into an appropriate character of the language (column 2, lines 12-15).

Nanjo et al. does not specifically disclose using the table to convert the index of textual data into a universal index. Converting the index of the textual data into a universal index would allow the textual data to be searched in any language.

It would have been obvious to one of ordinary skill in the art at the time of invention to further modify Nanjo et al., as modified by Holt et al., to convert the index of the textual data into a universal index so that the textual data could be searched in any language.

In regard to claim 20, Nanjo et al. discloses a query processor (search program 418) that transforms a query into corresponding semantic units (column 16, lines 1-19 and column 12, lines 43-48). The search program (418) also acts as a search engine

that searches the textual database based the semantic units corresponding to the search query (column 16, lines 40-43).

In regard to claim 21, Nanjo et al. discloses that word boundaries are automatically marked in the search query. If a separator character is found in the query, a key offset (KO) and key length (KL) value is stored delimits the boundaries of the word (column 16, lines 1-5).

11. Claims 9 and 17 are rejected under 35 U.S.C. 103(a) as being unpatentable over Nanjo et al in view of Makhoul et al. (U.S. Patent 5,933,525). Nanjo et al. discloses all the features of the instant claimed invention except identifying the type of textual data so that the step of transcribing is performed based on the type of textual data identified.

Makhoul et al. discloses an optical character recognition system that is language independent. The system can be used to recognize many languages (column 6, lines 34-43). This would suggest to one of ordinary skill in the art at the time of invention that the transcription step would depend on which type of language was recognized. Modifying Nanjo et al. to identify the type of textual data and transcribe based on that type of textual data would allow the use of orthographic rules for each particular language to be used in the recognition process, thereby minimizing the recognition search as taught by Makhoul et al. (column 6, lines 8-9 and 21-25).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify Nanjo et al. to include a step of identifying the type of textual data so

that the step of transcribing is performed based on the type of textual data identified in order to minimize the recognition search as taught by Makhoul et al.

In regard to claim 17, Nanjo et al. discloses all the features of the instant claimed invention, except a recognition system for transcribing textual data into corresponding semantic units wherein the recognition system comprises an OCR and an AHR.

Makhoul et al. discloses an OCR system that transcribes textual data into semantic units (words). Makhoul also discloses that the techniques used for character recognition (Hidden Markov Models) have been applied to AHR systems (on-line handwriting recognition systems) (column 1, lines 57-63). Modifying Nanjo et al. to include a recognition system comprising an OCR and an AHR would allow for alternative ways to transcribe textual data without typing.

It would have been obvious to one of ordinary skill in the art at the time of invention to modify Nanjo et al. to include a recognition system comprising an OCR and an AHR in order to provide alternative ways for transcribing textual data without typing.

Conclusion

12. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Brian L Albertalli whose telephone number is (703) 305-1817. The examiner can normally be reached on Monday - Friday, 8:30 AM - 5:00 PM.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Talivaldis Smits can be reached on (703) 305-3011. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

BLA 12/9/04

TĀLIVALDIS IVARS SMITS
PRIMARY EXAMINER